International Journal of Modern Physics A

Performance of an Operating High Energy Physics Data Grid: DØSAR-Grid

B. Abbott,[a] P. Baringer,[b] T. Bolton,[c] Z. Greenwood,[d] E. Gregores,[e] H. Kim,[f] C. Leangsuksun,[d]
D. Meyer,[f] N. Mondal,[g] S. Novaes,[e] B. Quinn,[h] H. Severini,*[a] P. Skubic,[a] J. Snow,[i]
M. Sosebee,[f] J. Yu [f]

[a] *University of Oklahoma, Norman, OK 73019, USA*
[b] *University of Kansas, Lawrence, KS 66045, USA*
[c] *Kansas State University, Manhattan, KS 66506, USA*
[d] *Louisiana Tech University, Ruston, LA 71272, USA*
[e] *Universidade Estadual Paulista, Sao Paulo, Brazil*
[f] *University of Texas at Arlington, Arlington, TX 76019, USA*
[g] *Tata Institute of Fundamental Research, Bombay, India*
[h] *The University of Mississippi, University, MS 38677, USA*
[i] *Langston University, Langston, OK 73050, USA*

Received (Day Month Year)
Revised (Day Month Year)

The DØ experiment at Fermilab's Tevatron will record several petabytes of data over the next five years in pursuing the goals of understanding nature and searching for the origin of mass. Computing resources required to analyze these data far exceed capabilities of any one institution. Moreover, the widely scattered geographical distribution of DØ collaborators poses further serious difficulties for optimal use of human and computing resources. These difficulties will exacerbate in future high energy physics experiments, like the LHC. The computing grid has long been recognized as a solution to these problems. This technology is being made a more immediate reality to end users in DØ by developing a grid in the DØ Southern Analysis Region (DØSAR), DØSAR-Grid, using all available resources within it and a home-grown local task manager, McFarm. We will present the architecture in which the DØSAR-Grid is implemented, the use of technology and the functionality of the grid, and the experience from operating the grid in simulation, reprocessing and data analyses for a currently running HEP experiment.

1. Introduction

Particle physicists employ high energy particle accelerators and complex detectors to probe ever smaller distance scales in an attempt to understand the nature and origin of the universe. The Tevatron proton-anti proton collider located in the Department of Energy's Fermi National Accelerator Laboratory,[1] Batavia, Illinois, currently operates at the "energy frontier" and has an opportunity of providing new

*Corresponding author. Tel.: +1-405-325-3961x36415; Fax: +1-405-325-7557; E-mail: severini@ou.edu.

discoveries through the DØ[2] and CDF[3] high energy physics (HEP) experiments. Future experiments, such as ATLAS[4] at the Large Hadron Collider (LHC)[5] in the European Organization for Nuclear Research (CERN)[6] and a proposed electron-positron linear collider (LC)[7] have the prospects of either building on discoveries made at Fermilab, or making the discoveries themselves if nature has placed these new processes beyond the energy reach of the Tevatron.

2. DØSAR

As part of the DØ experiment's effort to utilize grid technology for the expeditious analysis of data, twelve universities have formed a regional virtual organization (VO), the DØ Southern Analysis Region (DØSAR).[8] The centerpiece of DØSAR is a data and resource hub called a Regional Analysis Center (RAC),[9] constructed at the University of Texas at Arlington with the support of NSF MRI funds. Each DØSAR member institution constructs an Institutional Analysis Center (IAC), which acts as a gateway to other RACs and to the grid for the users within that institution. These IACs combine dedicated rack-mounted servers and personal desktop computers into a local physics analysis cluster. The data access system for DØ offline analyses is managed by a database and cataloging system called Sequential Data Access via Metadata, or SAM.[10] The MC_Runjob package[11] provides a low-level MC manager that coordinates data files used by and produced with the executables of the MC chain, via scripts produced for the job. This package is combined with MC task management software called McFarm[12], developed at UTA, for automated, highly efficient MC production.

3. DØSAR-Grid

In order to pursue full utilization of grid concepts, we are establishing an operational regional grid called DØSAR-Grid using all available resources, including personal desktop computers and large dedicated computer clusters. We have constructed and are operating DØSAR-Grid utilizing a framework called SAM-Grid[13] being developed at Fermilab. DØSAR-Grid will subsequently be made interoperable under other grid frameworks such as LCG,[14] TeraGrid,[15] Grid3,[16] and Open Science Grid.[17] Wherever possible, we will exploit existing software and technological advances made by these other grid projects. We plan to develop and implement tools to support easy and efficient user access to the grid and to ensure its robustness. Tools to transfer binary executables and libraries efficiently across the grid for environment-independent execution will be developed. DØSAR will implement the Grid for critical physics data analyses, while at the same time subjecting grid computing concepts to a stress test to its true conceptual level, down to personal desktop computers. Many of the proponents of this project are members of the LHC ATLAS[4] experiment and a future experiment under study by the American Linear Collider Physics Group,[18] and will apply the experience gained from DØSAR to these projects.

4. Conclusions

Experimental particle physics has always pushed information technology in directions that have had profound broader impact, with the most spectacular example being CERN's development of the World Wide Web, originally intended to facilitate communication among collaborating physicists spread around the globe. DØSAR's goal of efficiently using computing resources scattered across the south-central United States for tasks such as complex data reduction and simulation will produce grid computing solutions with broad applicability. This project will also likely affect how we disseminate and analyze data from future experiments at facilities such as the LHC. DØSAR will help improve the cyber-infrastructure within the region; five participating institutions are located in states that are traditionally under-funded for R&D activities. In addition to technology development, there will also be significant development of human resources in the region, delivering interdisciplinary training in physics and computer science. Information for the broader public will be disseminated on the web. Finally, DØSAR will realize the conceptual grid to the level of personal desktop computers, demonstrating the performance of a grid at its fullest level. The DØSAR Grid will be exploited in critical physics analyses of real and simulated data from the DØ experiment, and later from LHC and other future experiments. These data will serve to advance our understanding of nature at the smallest distance scales.

References

1. Fermilab National Accelerator Laboratory, http://www.fnal.gov/.
2. The DØ Experiment, http://www-d0.fnal.gov/.
3. The CDF Experiment, http://www-cdf.fnal.gov/.
4. The ATLAS (A Toroidal LHC ApparatuS) Experiment, http://www.cern.ch/Atlas/.
5. The Large Hadron Collider (LHC) project, http://www.cern.ch/lhc/.
6. European Organization for Nuclear Research (CERN), http://www.cern.ch/.
7. The Linear Collider Research and Development Working Group, http://www.hep.uiuc.edu/LCRD/.
8. The DØ Southern Analysis Region (DØSAR), http://www-hep.uta.edu/d0-sar/d0-sar.html.
9. I. Bertram, R. Brock, F. Filthaut, L. Lueking, P. Mattig, M. Narain, P. Lebrun, B. Thooris, J. Yu, C. Zeitnjitz, "Proposal for DØ Regional Analysis Centers," DØ Note #3984, unpublished (2002), http://www-d0.fnal.gov/computing/D0Race/d0rac-wg/d0rac-final.pdf.
10. The Sequential Data Access via Metadata (SAM) project, http://www-d0.fnal.gov/computing/sam/.
11. MC_Runjob Home Page, http://www-clued0.fnal.gov/runjob/.
12. McFarm Home Page, http://hepfm000.uta.edu/documentation.html.
13. The DØ Grid Group, http://www-d0.fnal.gov/computing/grid/.
14. The LHC Computing Grid Project, http://www.cern.ch/LCG/.
15. The TeraGrid, http://www.teragrid.org/.
16. The Grid3 project, http://www.ivdgl.org/grid2003/.
17. Open Science Grid, http://www.opensciencegrid.org/.
18. American Linear Collider Group (ALCPG), http://blueox.uoregon.edu/ lc/alcpg/.